

Using fuzzy logic for analysis and optimization of color-difference databases

Rafael Huertas^{1,*}, Daniel Arranz², Pedro Latorre-Carmona³ and Samuel Morillas⁴

Dpto. de Óptica, Universidad de Granada, Avda. de Fuentenueva s/n, 18071, Granada (Spain)

² Escuela de Ing. Informática, Universidad de Valladolid, P.º de Belén, 15, 47011 Valladolid (Spain)

³ Dpto. de Ingeniería Informática, Universidad de Burgos, Avda. Cantabria s/n, 09006. Burgos (Spain)

⁴ Instituto Universitario de Matemática Pura y Aplicada, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia (Spain)

Abstract

The precise measurement and specification of color, as well as the measurement of color differences between two color pairs of samples, are very important issues in Colorimetry, with applications in various fields such as automotive industry, textiles, agriculture, healthcare, etc. This relevance is especially pronounced in areas where color is not only an attribute but also adds significant value to the final product. Research carried out in this topic is mainly based on psychophysics experimentations on color differences perception. In this work, an exhaustive analysis of various color-difference databases has been carried out using data analysis techniques through fuzzy logic. The ultimate goal of this analysis has been to identify pairs of colors that are inconsistent compared to other pairs, in order to improve the quality and consistency of the initial databases. To achieve this, a methodology based on detecting discrepancies between visually perceived color differences and calculated color differences compared to the rest of the data has been implemented, using fuzzy logic methods. The result of this work is the identification, analysis, and elimination of pairs of colors considered inconsistent in various databases. This data cleansing significantly contributes to improving the quality of color-difference databases, which play a fundamental role in the development and evaluation of new color-difference formulas. In this way the results of these formulas align more precisely with the visual perception of color differences by the human visual system.

Keywords

Fuzzy logic, color differences, STRESS index.

1. Introduction

In the field of color differences, databases are essential, both for the development of new color difference formulas and checking their performance [1]. Each of the data in these databases is formed by the color coordinates of a pair of color stimuli and the perceived color difference between them, ΔV , which is measured through psychophysical experiments, and is the average of a considerable number of observers and/or several repetitions [2]. Thus, producing these databases is both time and resources consuming.

With the color coordinates of each of the stimuli in the pair, a color difference between them can be computed using different mathematical formulas, which can be just the Euclidean distance or other more sophisticated formulas. The calculated color difference is generally called ΔE , and it is desirable that ΔV and ΔE be as correlated as possible throughout the entire database.

* Corresponding author.

✉ rhuertas@ugr.es (Rafael Huertas); daniel.arranz.ort@gmail.com (Daniel Arranz); plcarmona@ubu.es (Pedro Latorre-Carmona); smorillas@mat.upv.es (Samuel Morillas)

ORCID 0000-0001-6606-0151 (Rafael Huertas); 0000-0001-6984-5173 (Pedro Latorre-Carmona); 0000-0001-9262-6139 (Samuel Morillas)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

The measurement of agreement between perceived and calculated color differences constitutes another field of study within Colorimetry. Different statistical measures have been proposed, and today the STRESS index [3] is one of the most used due to its best properties.

As we have mentioned, the consistency of databases is important for their reliability. We consider an inconsistency when two data, which must be close in a color space, and whose distance between the colors of the pair is similar, have very different perceived or computed color differences. The CIELAB color space is used in this work [4]. The closeness of the two data in the color space is computed by the Euclidean distance in CIELAB (ΔE_{ab}), while the distance between the two colors of each pair, ΔE , is computed by the CIEDE2000 (ΔE_{00}) [5], the current CIE/ISO-recommended color-difference formula. Other possibilities of color spaces and color difference formulas can be explored. To study the consistency of the database we will use fuzzy logic techniques, as described in the following previous works [2, 3], where two fuzzy logic rules are proposed.

The objective of this work is to analyze the effects of the different parameters and variables involved in the definition of this consistency analysis method. Preliminary results are shown here. A deeper analysis in the database is desirable to eventually provided cleaned version of the database.

2. Method

In [6, 7] authors have developed different fuzzy rules to identify the inconsistent data in a color-differences dataset. Here we analyze several experimental datasets through the application of this method. In addition, instead of a fixed threshold as in the published papers, in this work we study the effect of considering any value for the threshold in the application of the method.

2.1. Color-differences databases

In this work, a set of databases obtained from a series of psychophysical experiments will be considered. Specifically, four different databases with different structure and composition are studied. Each data of these databases consists of a pair of colors, specified by their color coordinates in a color space (i.e. CIELAB), the perceived color difference by a panel of observers and sometimes other significant data.

The considered databases are:

- COMData

The combined dataset was used in the development and performance testing of CIEDE2000 color-difference formula [5]. Besides, it has been used extensively for formula development and testing since its built, combining (COM) four different datasets: BFD-P, Witt, Leeds, and RIT-Dupont, as can be seen in [8]. Thus, it is an important and well-known database in the field of Colorimetry, used and studied in different works. It is one of the most extensive databases, with 3813 data points. The color differences, in CIELAB units, range from 0.04 to 18.21, average of 2.67 and standard deviation of 2.30. A complete description of this database can be found in [8].

- Pointer

Developed by researcher Michael Pointer and Geoffrey G. Attridge, at United Kingdom. It has 1308 color pairs distributed across 27 color centers. For each one, approximately 48 points, varying in L^* , a^* , and b^* have been tested. The color differences, in CIELAB unit, range from 0.88 to 26.21, average of 8.91 and standard deviation of 4.47. Some details can be found in [9].

- LCAM-WDC

Developed by researcher Michal Vik in the Faculty of Textile Engineering of the University of Liberec, in Czech Republic. It has 284 color pairs around 9 color centers. The color differences, in CIELAB units, range from 0.05 to 13.72, average of 2.53 and standard deviation of 2.64.

- RIT-Dupont-Ind

The complete name of the database is RIT-DuPont Supra-Threshold Color-Tolerance Individual Color-Difference Pair. This is the complete database of an extensive experiment conducted at the Rochester Institute of Technology (RIT) in the USA. These original judgments were transformed, by probit analysis, to 156 color pairs in the so-called RIT-DuPont dataset, which is one of the four included in the COMData. Thus, it is the extensive original dataset, which would be worthy analyses independently of the synthesized version inside COMDATA. It has 828 color pairs, distributed across 19 color centers. Each experimental data was assessed by a panel of 50 observers, consequently this database is considered as one of the most reliable. The color differences, in CIELAB units, range from 0.03 to 5.42, average of 1.50 and standard deviation of 0.64. A complete description of this database can be found in [10].

2.2. Standardized Residual Sum of Squares (STRESS)

The figure of merit known as STRESS[3] (Standardized Residual Sum of Squares) is used in a not insignificant number of different contexts [11] to check the accuracy of a model, i.e. color difference formulae. In this work, STRESS is considered to evaluate the agreement between the set of visual data (ΔV) and the corresponding color differences (ΔE) computed with different formulae. The lower the value of STRESS, the better agreement between the two sets, ΔV and ΔE .

Assuming that a color difference formula is accurate enough, STRESS can be considered as a measurement of the consistency of certain dataset. Therefore, removing inconsistent data will imply a reduction in the value of STRESS.

As shown in the previous paper[3], to compute STRESS, a previous factor (scaling factor, F) must be computed, to scale either ΔV or ΔE , in such a way that both are in the same scale. Please, note that in Colorimetry it is not important that the scales of ΔV and ΔE could be different, while ΔV and ΔE are correlated. In [3] it can also be seen that the STRESS index can be computed by three different equations, each one of them having a different scaling factor, called F_1 (applied to ΔV), F_2 (applied to ΔE), and F_3 (applied to ΔV).

2.3. Fuzzy rules

Each data of the dataset is compared with the rest of the data to inquire into its consistency. Thus, let us consider two data (i.e. two color-pairs), which are close in the color space. The two data are considered inconsistent if any of these cases applied:

- The two pairs are very close in color space and have similar ΔE values but very different ΔV values.
- The two pairs are very close in color space and have similar ΔV values but very different ΔE values.

Thus, the important variables are the distance between the color pairs and the comparison between the perceived and computed color differences. All this reasoning is formulated by means of vague linguistic terms; thus, fuzzy logic can be used for its numerical representation. The two former cases, were implemented in two fuzzy rules, known as *Fuzzy Rule 1* (FR1) and *Fuzzy Rule 2* (FR2), which detect inconsistencies in a color-differences dataset. The details can

be found in [7]. Please, note that in the definition of these fuzzy rules, 1 and 2 are identical exchanging ΔE and ΔV . *Fuzzy Rule 1* focuses on evaluating inconsistency based mainly on differences in perceived color difference (ΔV) between data pairs, while *Fuzzy Rule 2* does this in computed color differences (ΔE). Usually, the same data inconsistent with a fuzzy rule are also with the other, but this is not always the case.

The first condition, that the two pairs are close, is related to the distance in the corresponding color space. In fuzzy logic, for an inference rule to compute certainty of the rule consequent it needs the antecedent to have not null certainty. Otherwise, no inference is carried out. The certainty of the antecedent for two color pairs is called *fuzzy degree of neighborhood* and it is defined for each pair considering the rest of pairs in the database. It takes into account the distance and the similarity of ΔV between the two data.

Therefore, these fuzzy rules are applied for each color pair with respect to all other color pairs in the set to find out whether there exist inconsistencies. Eventually, all possible color pairs are compared. It may happen that for some color pairs there is no other enough similar color pairs to be compared to. This should not be confused with color pairs for which there indeed exist similar color pairs to be compared to but no inconsistency is observed for them. To distinguish between these two cases, we propose to compute what is called the *number of fuzzy neighbors* for each color pair. This is computed as the sum of the certainty of the antecedent of the fuzzy rule when computed for all other data. For the result of the fuzzy rule for a color pair to be considered representative, its *number of fuzzy neighbors* must be at least equal to 1. Otherwise, the procedure considers that there is not enough evidence to make conclusions about the color pair.

For the application of any of defined fuzzy rules it is necessary that ΔV and ΔE are on the same scale, so a normalization of ΔV is carried out following the procedure in [3]. There are two possibilities to scale ΔV , F_1 or F_3 . Similar results are obtained for both, but F_3 is preferable, as shown in a previous work [12], and it is considered in this study.

Finally, these fuzzy rules provide a number, I_{ij} in the interval [0,1], representing the degree of inconsistency between the two data. Higher values (its maximum value being 1) indicate greater inconsistency between the couple of color pairs. We will analyze and compare the results after the application of each of the two fuzzy rules, and the application of the combination of both, i.e. joining the inconsistency data from the application of both fuzzy rules, with means the logical operator FR1 or FR2.

Once the degree of inconsistency is computed, a threshold value must be defined in order to consider an inconsistency or not. The threshold represents the measure of how similar two pairs of colors need to be to be considered consistent. When the threshold is low, such as 0.05, a very high similarity value between the pairs of colors is required for them to be considered consistent. Therefore, it is more likely that inconsistent pairs will be detected. Authors in [7] considered 0.5 as a reasonable value for a specific database. In this work the results have been computed for thresholds between 0 and 1 with a step of 0.05.

In the next step, in order to clean the database, it is decided which of the two color-pairs must be removed when the degree of inconsistency is higher than the threshold. The decision is to remove, between the two pairs, the one with higher difference between ΔV and ΔE .

3. Results and conclusions

Firstly, the distribution of experimental data based on the number of fuzzy neighbors is analyzed for the different databases.

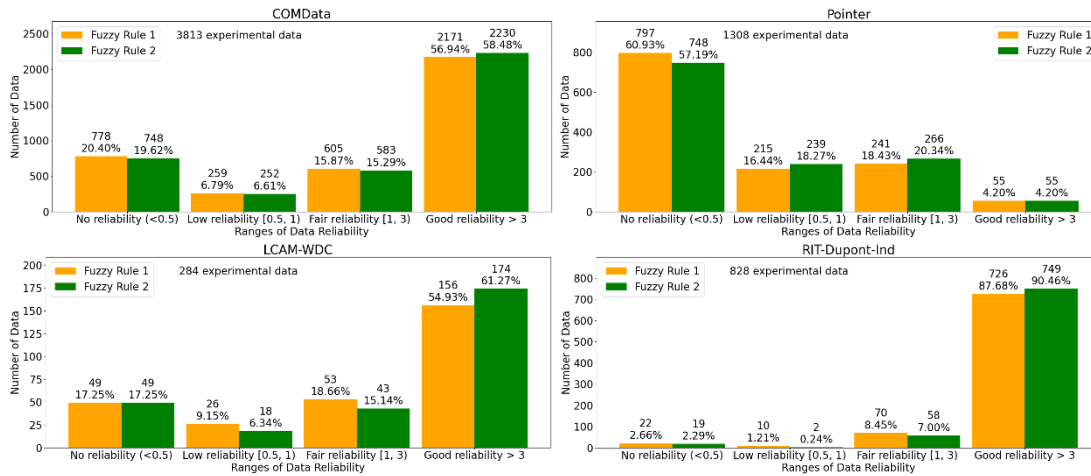


Figure 1: Number of fuzzy neighbors for COM (top left), Pointer (top right), LCAM-WDC (bottom left) and Rit-Dupont-Ind. (bottom right).

Fig. 1 shows the clustering of the experimental data from the four databases based on the number of fuzzy neighbors (the total fuzzy degree of neighborhood sum). When the number of fuzzy neighbors is lower than 0.5, there is a significant lack of surrounding information to determine the quality of an experimental datum. In this scenario, the evaluation can be highly uncertain due to the scarcity of nearby or similar reference points. It is desirable a number of fuzzy neighbors at least equal to 1, or higher, to make informed decisions about their consistency of the data. In Fig. 1 we can see that most of the data in Poiter database has too low number of fuzzy neighbors, while in COM, RIT-Dupont-Ind and LCAM-WDC databases most of the data has enough fuzzy neighbors to do the comparison.

Once we know the number of fuzzy neighbors, the degree of inconsistency can be worked out between each combination of two data in each database. Fig. 2 shows the number of inconsistent pairs as a function of the threshold. These plots provide valuable insight into how the choice of threshold affects the identification of inconsistent experimental data.

As expected, the plots show an increase in the number of inconsistent pairs as the threshold decreases and vice versa. The number of inconsistent pairs against the thresholds has a tendency between linear and exponential depending on the database. In some databases, the number of inconsistent pairs reaches zero as the threshold increases, suggesting that the data is highly consistent within that threshold range. However, this point can vary depending on the specific dataset, highlighting the importance of customizing the threshold according to the context and requirements of the study. Accordingly, initially the most consistent database is the COMData. In general, some differences can be found between the two fuzzy rules, also depending on the database. A particular case is RIT-Dupont-Ind, where only RF2 obtains inconsistencies. The accuracy of this database makes that no inconsistencies at all are found with FR1, and only few with FR2.

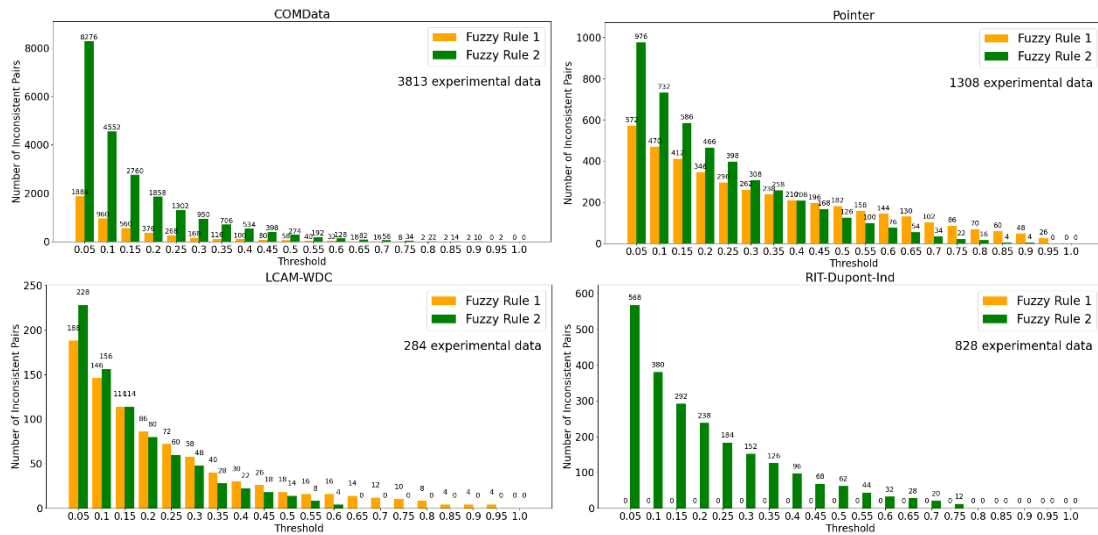


Figure 2: Number of inconsistent pairs as a function of the threshold for COMData (upper left), Pointer (upper right), LCAM-WDC (bottom left) and Rit-Dupont-Ind. (bottom right).

The next step, after identifying the inconsistencies, is cleaning and optimizing the database. To check the efficiency of the process, the STRESS value is compared with the uncleaned/initial version. This methodology can give us the criterion to choose the threshold: A threshold can be admitted if/while the STRESS improves considerably when removing the corresponding inconsistent data. Obviously, the quality of the database will improve as more inconsistent data are removed. However, there is a critical point in this process. There comes a point when the experimental data remaining in the database after removal are equally consistent compared to those that were removed. At this point, the STRESS value stops improving and remains practically constant. This approach seeks to find a balance between improving the quality of the database by removing inconsistent data and preserving the original nature of the database. If too many data are removed, the database could lose its original representation and become a smaller, but highly consistent, set. Therefore, this criterion aims to find the balance at which the quality of the database is maximized without losing its initial integrity.

Fig. 3 shows the results for the COMData. The combination of the two fuzzy rules is the most effective approach, removing more data and achieving the lowest STRESS values. This result will be general for any database. In this case, thresholds between 0.55 and 0.50 get the lowest improvement in STRESS, being these values the desired balance.

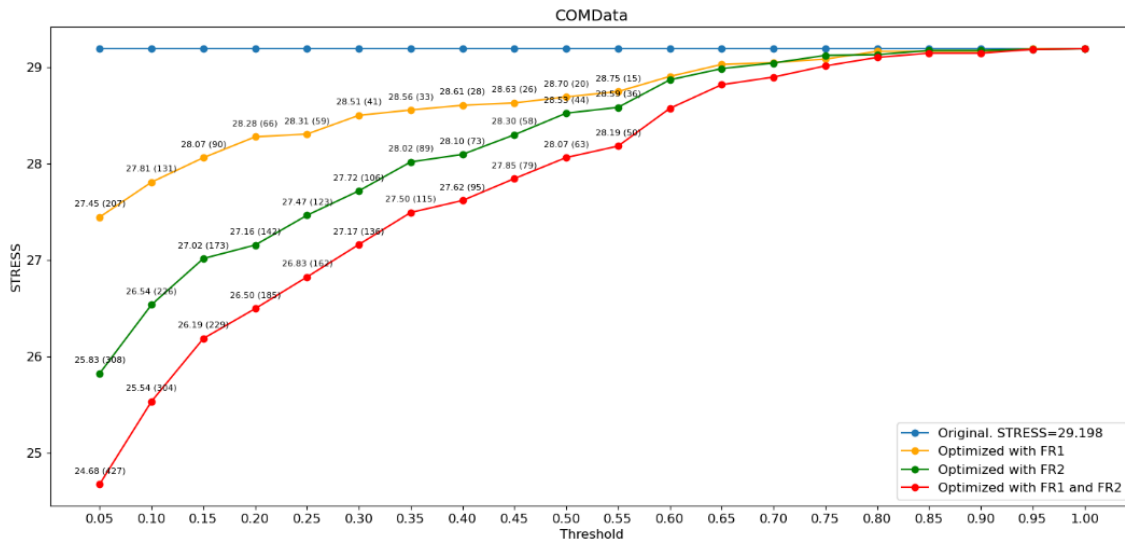


Figure 3: STRESS values for the COMData after removing the inconsistent pairs determined by the FR 1, FR2 and the combination, as a function of the threshold. For each threshold the STRESS value and the number of removed data, in brackets, are given.

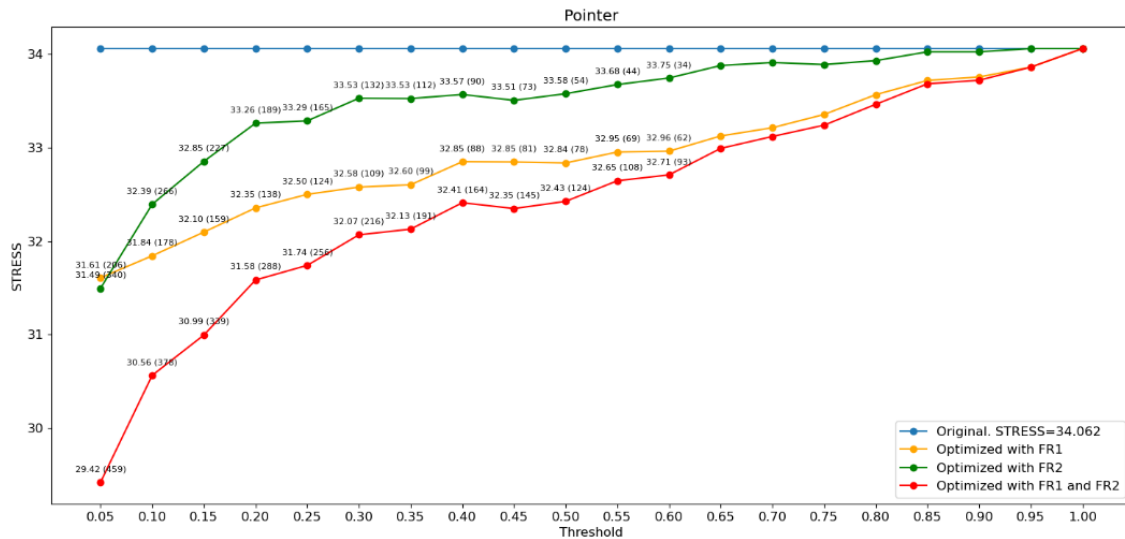


Figure 4: STRESS values for the Pointer after removing the inconsistent pairs determined by the FR 1, FR2 and the combination, as a function of the threshold. For each threshold the STRESS value and the number of removed data, in brackets, are given.

In the case of Pointer, similar results are obtained, but FR1 is better than FR2. In this case, a threshold of 0.5 is a suitable candidate. This means that by removing experimental data below this threshold, a substantial improvement in the STRESS level is achieved, as it can be seen in Fig. 4. For FR2, less than 4.2% of the data, 6.0% for FR1, and for the combination less than 9.5% of the data would be removed.

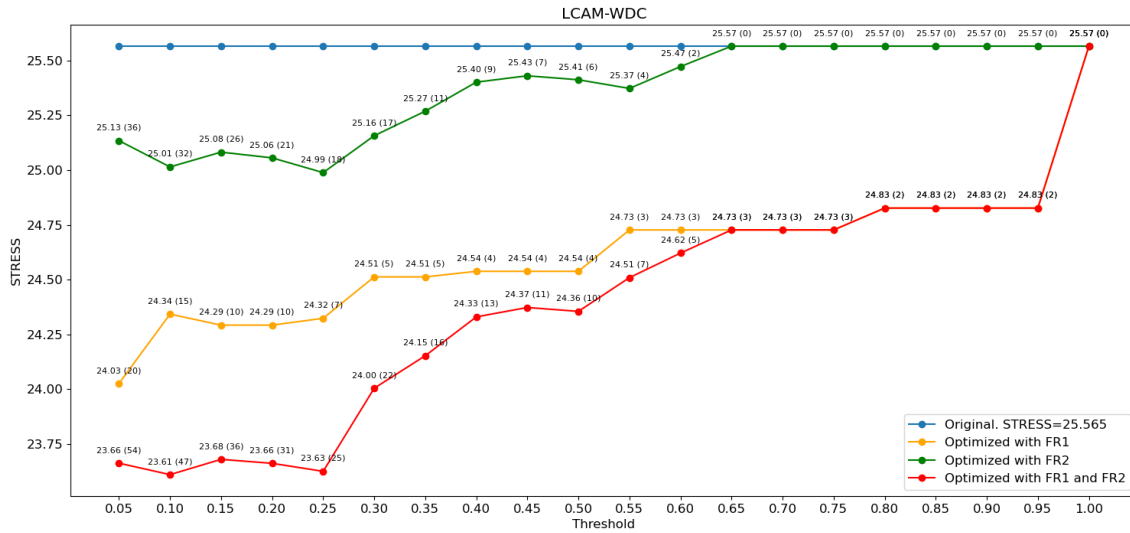


Figure 5: STRESS values for the LCAM-WDC after removing the inconsistent pairs determined by the FR 1, FR2 and the combination, as a function of the threshold. For each threshold the STRESS value and the number of removed data, in brackets, are given.

For the LCAM-WDC database, as it is shown in Fig. 5, the results are not so linear. Three different values of the threshold could be considered. The most conservative one is 0.95, achieving an important improvement in the STRESS value, by removing only 2 data (0.7%). This means that these 2 data are highly inconsistent in this dataset and could be a typo. Threshold must decrease until 0.75 to find more inconsistent data (3 data). Again, a good balance seems to be a threshold of 0.5, which removes 10 data (3.5%) and reduces the STRESS value in 1.2 points.

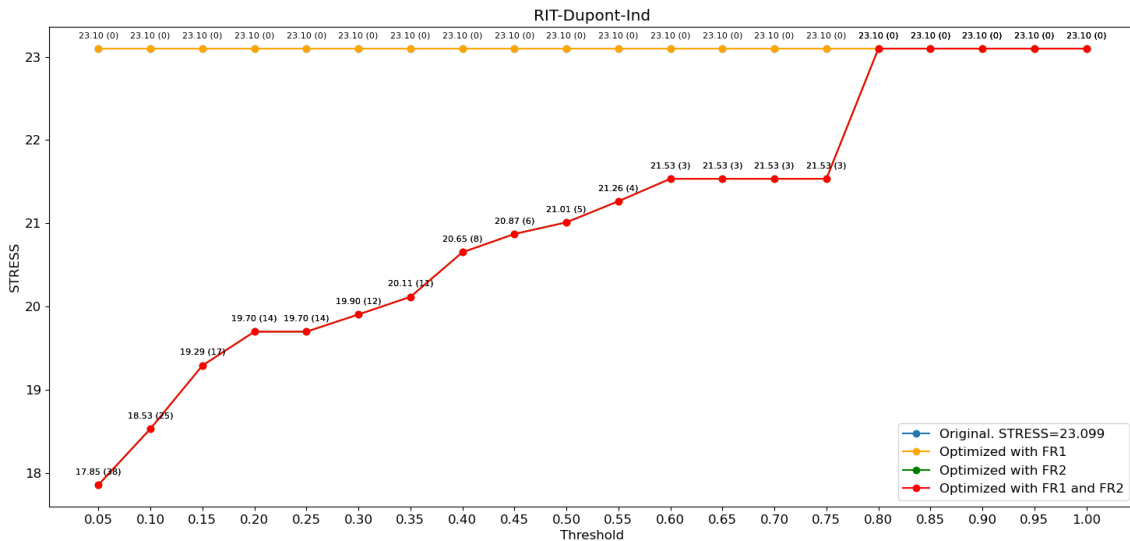


Figure 6: STRESS values for the RIT-Dupont-Ind after removing the inconsistent pairs determined by the FR1, FR2 and the combination of both, as a function of the threshold. For each threshold the STRESS value and the number of removed data, in brackets, are given.

Fig. 6 shows the result of the RIT-Dupont-Ind database, which is a special case, as commented above. FR 1 is no able to detect inconsistencies, thus FR2 and the combination give the same results. This is a quite consistent dataset, since inconsistencies only appear for threshold values lower than 0.80 (3 data, 0.36%) and 0.55 (4 data, 0.48%). Even with the extremely low threshold value of 0.1, only 25 data (3%) can be considered as inconsistent.

As a result of this analysis, we have observed that applying fuzzy logic to color-differences databases provides substantial information about the consistency of them. Also, it is possible to identify and remove the inconsistencies up to a balance between improvement and reducing the number of data is achieved. These results indicate that optimal threshold values vary significantly depending on the database and the applied rule. Finding a balance is crucial, as overly low thresholds can lead to excessive removal of useful data, while overly high thresholds may fail to detect inconsistencies.

References

- [1] M. Melgosa, "Request for existing experimental datasets on color differences," Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur **32**, 159 (2007).
- [2] E. Kirchner, N. Dekker, M. Lucassen, L. Njo, I. van der Lans, P. Koeckhoven, P. Urban, and R. Huertas, "How color difference formulas depend on reference pairs in the underlying constant stimuli experiment," JOSA A **32**, 2373-2383 (2015).
- [3] P. A. Garcia, R. Huertas, M. Melgosa, and G. Cui, "Measurement of the relationship between perceived and computed color differences," JOSA A **24**, 1823-1829 (2007).
- [4] C. Colorimetry, "CIE Publication No. 15.2, Central Bureau of the CIE, Vienna, 1986," The commonly used data on color matching functions is available at the CIE web site at <http://www.cie.co.at>.
- [5] ISO/CIE 11664-6: 2014 (Formerly CIE S 0146/E: 2013), "Colorimetry—Part 6: CIEDE2000 Colour-Difference Formula," (2014).
- [6] S. Morillas, L. Gómez-Robledo, R. Huertas, and M. Melgosa, "Fuzzy analysis for detection of inconsistent data in experimental datasets employed at the development of the CIEDE2000 colour-difference formula," Journal of Modern Optics **56**, 1447-1456 (2009).
- [7] S. Morillas, L. Gómez-Robledo, R. Huertas, and M. Melgosa, "Method to determine the degrees of consistency in experimental datasets of perceptual color differences," JOSA A **33**, 2289-2296 (2016).
- [8] M. Melgosa, R. Huertas, and R. S. Berns, "Performance of recent advanced color-difference formulas using the standardized residual sum of squares index," JOSA A **25**, 1828-1834 (2008).
- [9] M. R. Pointer and G. G. Attridge, "Large colour differences in colour reproduction—the relationship between print acceptability and colour difference," The Journal of Photographic Science **44**, 155-164 (1996).
- [10] R. S. Berns and B. Hou, "RIT-DuPont supra-threshold color-tolerance individual color-difference pair dataset," Color Research & Application **35**, 274-283 (2010).
- [11] P. Latorre-Carmona, R. Huertas, M. Pedersen, and S. Morillas, "Proposal of a new fidelity measure between computed image quality and observers quality scores accounting for scores variability," Journal of Visual Communication and Image Representation **90**, 103704 (2023).
- [12] D. Arranz, R. Huertas, P. Latorre-Carmona, and S. Morillas, "Optimization of a color difference database by fuzzy logic," in *XIV Reunion Nacional de Optica*, Anonymous (Accepted).